Análisis de regresión lineal simple
con ordenada al origen



Ing. Byron Humberto González Ramírez
Profesor de Estadística Aplicada
Facultad de Ciencias Económicas y Empresariales
Universidad Rafael Landívar
bhgonzalez@url.edu.gt
http://www.byrong.tk
Publicación: septiembre 2006

1. Presentación

Para el desarrollo de este tema se asume que se ha estudiado previamente el cálculo de los dos estimadores de un modelo de regresión lineal simple. Un detalle de estos cálculos puede verse en http://www.stat.tamu.edu/stat30x/notes/node155.html#SECTION00120000000000000000 .

En algunos casos en los que se necesita ajustar modelos de regresión, puede ser posible que no se requiera calcular el intercepto. Esta condición se debe a la relación entre las variables analizadas.

Debemos recordar que el intercepto nos indica las unidades de la variable Y cuando X es cero. Pero en estos casos resulta que cuando X es cero Y también lo es.

En forma natural es posible ver esta condición al analizar la relación entre las variables involucradas. Por ejemplo si relacionamos la población de una plaga con el daño que causa. Es claro que si la población es cero, el daño también será cero.

Otros ejemplos de parejas de variables que no requieren el cálculo del valor del intercepto son: nivel de contaminación y casos de enfermedad, número de fallas en un equipo y tiempo de espera, etc.

Enseguida presentamos las ecuaciones mínimos cuadráticas para el cálculo de un modelo de regresión sin incluir el intercepto. También se acompañan los resultados provistos por SAS y SPSS V.12.0

2. Ejemplo de aplicación

Presentamos un ejemplo desarrollado por Martínez G., A; Castillo M., A. (1987).

Por algunas razones ocurre que en las zonas cañeras la programación de los cortes y el transporte de la caña no están bien equilibrados, ocasionándose varios problemas. Uno de los más frecuentes es el de la caña que tiene que abandonarse en el campo hasta que se consiguen los medios para llevarla al Ingenio para su procesamiento. Cuando esto ocurre después de varios días de cosechada la caña, se producen pérdidas cuyo monto puede a llegar a ser considerable.

En el ejemplo siguiente se tomaron muestras por un lapso de dos semanas a diferentes variedades de caña con el objeto de determinar algunas de las características químicas y físicas de la caña cosechada y abandonada en el lugar de corte. En la tabla 1 se presentan las pérdidas medias en porcentaje de peso de caña, estimadas diariamente a partir del día de cosecha. Se desea investigar el efecto del tiempo en días después de cosechada la caña (X), sobre el porcentaje de pérdida en peso.

Tabla 1: Deterioro de la caña en relación a los días transcurridos posterior a la cosecha.

Días después de la cosecha (x)	% Pérdida en peso (y)	Días después de la cosecha (x)	% Pérdida en peso (y)
0	0.00	8	12.07
1	2.16	9	14.4
2	4.4	10	15.28
3	6.08	11	16.62
4	7.31	12	18.77
5	8.8	13	18.6
6	9.9	14	20.25
7	10.94		

Observe que el día de la cosecha el porcentaje de pérdida en peso de la caña, es indudablemente igual a cero (0). Si ambos conjuntos de datos se grafican, los puntos siguen una tendencia rectilínea forzada a pasar por el origen, puesto que el día de la cosecha $x = 0$ *y* $Y = 0$. Esta condición nos hace pensar que un modelo de regresión sin intercepto es apropiado para representar los resultados de esta investigación. Vea la figura 1..

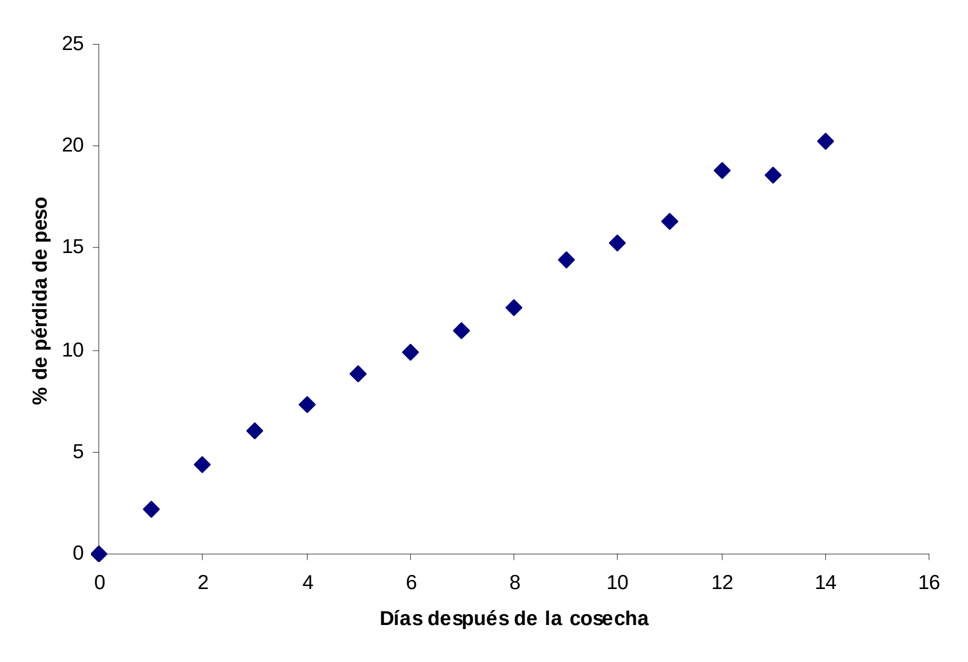


Figura 1: Diagrama de dispersión entre los días posteriores a la cosecha y el porcentaje de pérdida de peso de la caña

Cuando el modelo de regresión simple no contiene una ordenada al origen, el término ***b*** $\boldsymbol{b_o}$ de la expresión es ignorado, y el modelo se reduce a la forma:

$$Y_i = \beta X_i + e_i$$

donde,

$$\beta = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Con relación a la hipótesis, ß = 0, la suma de cuadrados debida a la regresión, es ahora:

$$SC\ Reg = \beta * (\sum_{i=1}^{n} X_i Y_i)$$

No hay corrección debida a la media y la tabla de análisis de varianza toma la estructura de la tabla 2, donde la SCe, se calcula por diferencia, restando de la suma de cuadrados total sin corregir.

$$SC_{Total} = \sum_{i=1}^{n} Y_i^{2}$$

Tabla 2: Análisis de varianza de la regresión al origen.

FV	GL	SC	CM	Fc
Regresión	1	SC_{Reg}	SC_{Reg} / 1	CM_{Reg} / CM_e
Error	n - 1	SC_{Tot} - SC_{Reg}	SC_{ee} / (n-1)	
Total	n	SC_{Total}		

Operando adecuadamente, para el conjunto de datos presentados, se obtiene el modelo siguiente:

Y = 1.5278 X ; es decir

% de pérdida de peso = 1.5278 días después de la cosecha

El análisis de varianza para éste modelo se presenta en la tabla 3, donde puede apreciarse el rechazo de la Ho, con un nivel de significancia de 1%.

Tabla 3: Resumen del análisis de varianza de la regresión al origen.

FV	GL	SC	CM	Fc	$Ftab_{0.01}$
Regresión	1	2369.22	2369.255	2704*	9.07
Error	13	11.3891	0.8761		
Total	14	2380.6448			

Concluimos que el modelo encontrado es adecuado para explicar la relación entre las variables analizadas. Adicionalmente, tenemos que por cada día que se deja la caña cosechada tirada en el campo hay una pérdida en porcentaje de peso, de 1.5278.

3. Procesamiento electrónico

3.1 Regresión al origen con SAS

A continuación se presenta el programa de ingreso, y los resultados obtenidos al emplear el *Sistema de Análisis Estadístico (SAS)* para ajustar un modelo con ordenada al origen.

Programa de Ingreso:

```
OPTIONS NODATE;
DATA regres;
INPUT T P;
LABEL T="Tiempo post-cosecha en días"
        P="Porcentaje de pérdida en peso";
CARDS;
1   2.16
2   4.4
3   6.08
4   7.31
5   8.8
6   9.9
7   10.94
8   12.07
9   14.4
10  15.28
11  16.62
12  18.77
13  18.6
14  20.25
;
PROC REG;
        MODEL P=T / NOINT;
RUN;
```

El ajuste el modelo de regresión se logra a través del procedimiento REG de SAS. Donde la instrucción MODEL requiere la opción NOINT, para procesar la regresión con ordenada al origen. La computadora envía como salida, la hoja de la tabla 4.

Tabla 4: Resultados del análisis de regresión al origen empleando SAS.

```
Model: MODEL1
Dependent Variable: P  "Porcentaje de pérdida en peso"

                             Analysis of Variance

                                Sum of          Mean
       Source          DF      Squares        Square        F Value        Prob>F

       Model            1   2369.25571    2369.25571       2704.371        0.0001
       Error           13     11.38909       0.87608
       U Total         14   2380.64480

           Root MSE        0.93599     R-square        0.9952
           Dep Mean       11.82714     Adj R-sq        0.9948
           C.V.            7.91394

                         Parameter        Standard     T for H0:
     Variable   DF        Estimate           Error   Parameter=0     Prob > |T|
     T           1        1.527823      0.02937919        52.004         0.0001
```

3.2 Regresión al origen con SPSS

Describimos el procedimiento que se sigue en SPSS para calcular nuestro modelo de regresión sin calcular el valor del intercepto.

En principio nos dirigimos al menú *Analizar / Regresión* como se muestra en la figura 2. Esto nos permitirá acceder a la ventana donde se define la posición de las variables analizadas para calcular el modelo.

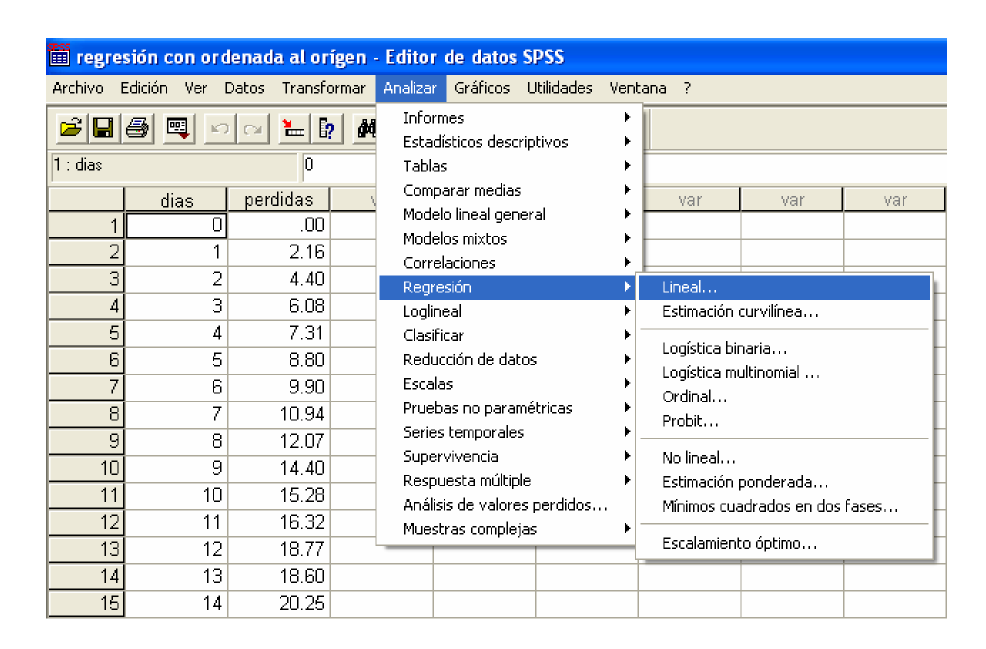


Figura 2: Menú para solicitar regresión en SPSS

Movemos la variable *Porcentaje de pérdida* a la posición de la variable dependiente y la variable *Días posterior a la cosecha* al lugar de las variables independientes.

Adicionalmente hacemos clic en el botón opciones para poder modificar el cálculo de las sumas de cuadrados para nuestro modelo. Esto se muestra en la figura 3.

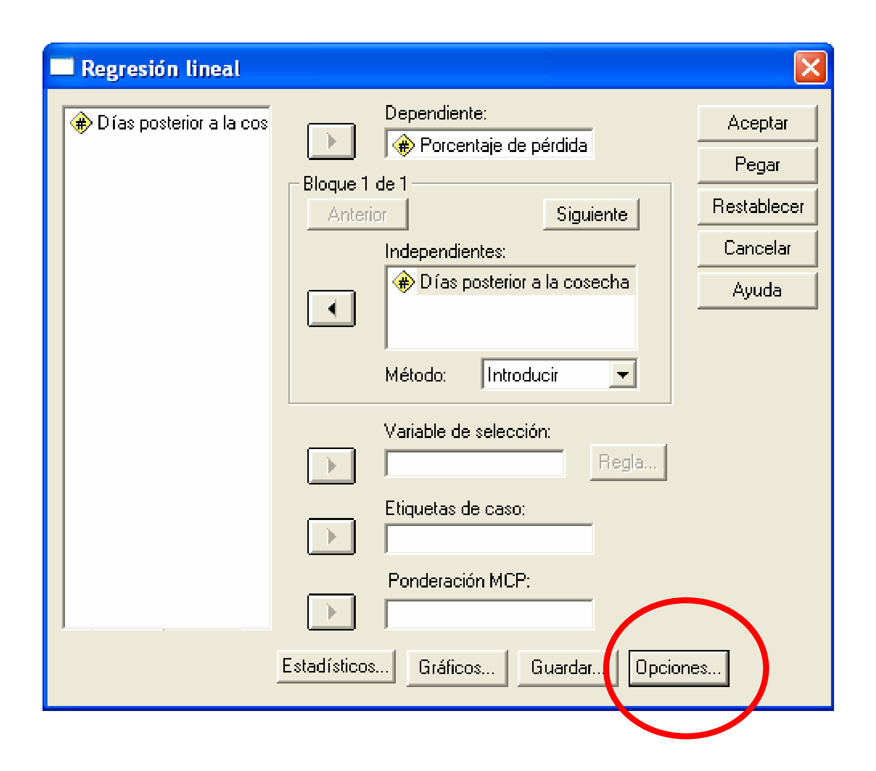


Figura 3: Definición de variables para el modelo y acceder a opciones

Aparecerá en pantalla la posibilidad de modificar algunas opciones para el cálculo de nuestro modelo. De las opciones disponibles *desmarcamos* la casilla correspondiente a *Incluir constante en la ecuación,* para que el programa no calcule el valor del intercepto.

Si quisiéramos también podríamos modificar el nivel de significancia para probar las hipótesis acerca de la validez de nuestro modelo. Este cambio afectaría las regiones de rechazo para el análisis de varianza y la prueba de “t” para la pendiente del modelo. Esto se muestra en la figura 4.

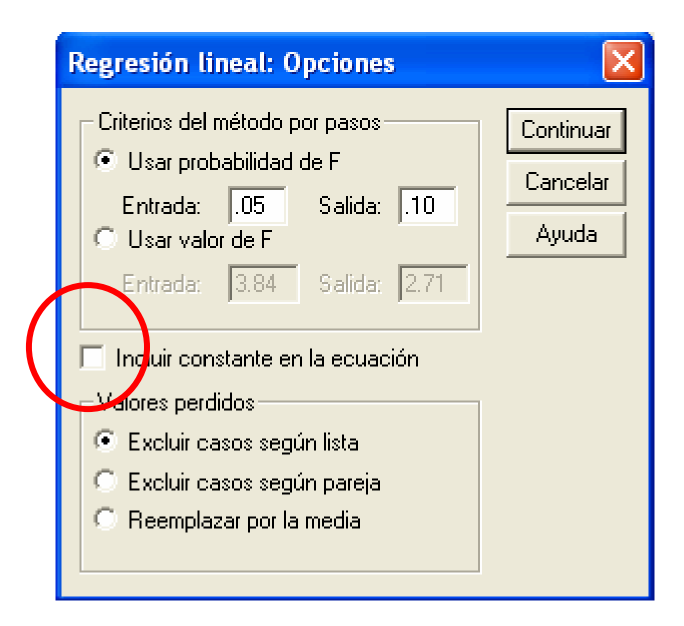
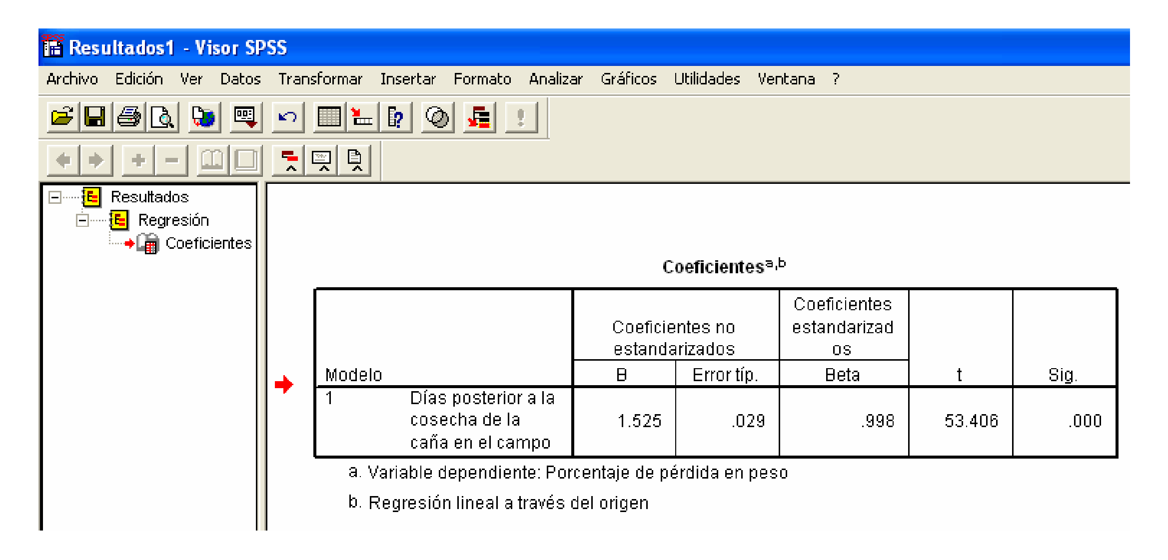


Figura 4: Desmarcar la casillas de verificación para impedir el cálculo de la constante en nuestro modelo

Una vez modificadas las opciones que nos interesan presionamos el botón continuar. Finalmente hacemos clic en el botón *Aceptar* para obtener los resultados. En nuestro caso mostramos en la figura 5 parte de las salidas generadas por SPSS.

Resaltamos que será necesario analizar los valores de los coeficientes de determinación, de correlación, y el gráfico de residuales con la variable dependiente para poder evaluar a fondo la validez del modelo calculado.



Coeficientes[a,b]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	Días posterior a la cosecha de la caña en el campo	1.525	.029	.998	53.406	.000

a. Variable dependiente: Porcentaje de pérdida en peso

b. Regresión lineal a través del origen

Figura 5: Modelo calculado por SPSS con ordenada al origen

4. Referencias

Martínez G., A; Castillo M., A. 1987. Teoría de la regresión con aplicaciones. Centro de Estadística y Cálculo. Colegio de Posgraduados, Chapingo. México. pp. 118-127. ISBN 968-839-04.

Gonçalvez C., F. 2002. Estatística. Universidade Estadual de Londrina. Brasil 304p. ISBN 85-7216-328-X.

Levin D. et al. 2000. Estatística. Teoria y Aplicações usando Microsoft Excel em português. Prentice Hall. pp. 513-578.